Probabilistic Dynamic Logic of Phenomena and Cognition

Evgenii Vityaev, Boris Kovalerchuk, Leonid Perlovsky, Stanislav Smerdov

Abstract—**The purpose of this paper is to develop further the main concepts of Phenomena Dynamic Logic (P-DL) and Cognitive Dynamic Logic (C-DL), presented in the previous paper. The specific character of these logics is in matching vagueness or fuzziness of similarity measures to the uncertainty of models. These logics are based on the following fundamental notions:** *generality relation, uncertainty relation, simplicity relation, similarity maximization problem with empirical content and enhancement (learning) operator.* **We develop these notions in terms of logic and probability and developed a Probabilistic Dynamic Logic of Phenomena and Cognition (P-DL-PC) that relates to the scope of probabilistic models of brain. In our research the effectiveness of suggested formalization is demonstrated by approximation of the expert model of breast cancer diagnostic decisions. The P-DL-PC logic was previously successfully applied to solving many practical tasks and also for modelling of some cognitive processes.**

I. Introduction

In the paper [1] there was introduced a Phenomena Dynamic Logic (P-DL) and Cognitive Dynamic Logic (C-DL) as a generalization of the Dynamic Logic and Neural Modelling Fields theory (NMF) introduced in the previous papers [2], [3]. Logics P-DL, C-DL provide the most general description of Dynamic Logic in the following fundamental notions *generality relation, uncertainty relation, simplicity relation, similarity maximization problem with empirical content and enhancement (learning) operator*. This generalization provide interpretation of P-DL, C-DL logics in the frame of other approaches.

In this paper we interpret logics P-DL, C-DL in terms of logic and probability: uncertainty we interpret as probability, while the process of learning as a semantic probabilistic inference [4], [9], [6], [5]. We also interpret mentioned fundamental notions. The resulting Probabilistic Dynamic Logic of Phenomena and Cognition (P-DL-PC) belong to the scope of the probabilistic models of brain [19], [20]. Thus, through logics P-DL, C-DL we extend the interpretation of Dynamic Logic and Neural Modelling Fields theory to probabilistic models of brain. The P-DL-PC logic as probabilistic model of brain was previously applied to modelling of some cognitive process [7], [8], [9], [21]. The effectiveness of P-DL-PC logic

Evgenii Vityaev is with the Department of Mathematical Logic, Sobolev Institute of Mathematics of the Russian Academy of Sciences and with the Department of Discrete mathematics and Informatics of the Novosibirsk State University, 630090, Novosibirsk, Russia, email: vityaev@math.nsc.ru

Boris Kovalerchuk is with the Department of Computer Science, Central Washington University, Ellensburg, WA 98926-7520, e-mail: borisk@cwu.edu

Leonid Perlovsky is with the Harvard University and the Air Force Research Laboratory, Sensors Directorate, Hanscom AFB, leonid@seas.harvard.edu

Stanislav Smerdov, Novosibirsk State University, Sobolev Institute of Mathematics of the Russian Academy of Sciences, 630090, Novosibirsk, Russia, email: netid@ya.ru

demonstrated in this paper by approximation of the expert model of breast cancer diagnostic decisions.

II. Universal productions. Data for prediction

In our study learning models will be generated as sets of *universal productions* (*u-productions*), which are introduced in this section. Note that every set of *universal formulas* is logically equivalent to a certain set of u-productions.

Consider a fixed first-order language $\mathfrak{L}$ in a countable signature. Hereafter denote $\mathbf{A}_\mathfrak{L}$ the set of all atoms; $\mathbf{L}_\mathfrak{L}$ – the set of all literals; $\mathbf{S}_\mathfrak{L}^0$ – the set of ground sentences. The set of ground atoms and the set of ground literals are denoted $\mathbf{A}_\mathfrak{L}^0 \rightleftharpoons \mathbf{A}_\mathfrak{L} \cap \mathbf{S}_\mathfrak{L}^0$ and $\mathbf{L}_\mathfrak{L}^0 \rightleftharpoons \mathbf{L}_\mathfrak{L} \cap \mathbf{S}_\mathfrak{L}^0$ correspondingly. Following examples of atoms and literals are given in the section VIII for the task of approximation of the expert model of breast cancer diagnostic decisions: 'number of calcifications per cm^3 less than 20', 'volume of calcifications in cm^3 not less or equal to 5', 'total number of calcifications more than 30 and etc.

Let Θ be the set of all substitutions and $\Theta^0 \subseteq \Theta$ the set of ground substitutions, that are mappings variables to ground terms. All necessary notions from model theory and logic programming are elementary and can be easily found in books [12], [13], [14].

Definition. A record of the type

$$\mathrm{R} \rightleftharpoons \tilde{\forall}(\mathrm{A}_1 \wedge \cdots \wedge \mathrm{A}_m \Leftarrow \mathrm{B}_1 \wedge \cdots \wedge \mathrm{B}_n),$$

where $\mathrm{A}_1, \cdots \mathrm{A}_m, \mathrm{B}_1, \cdots, \mathrm{B}_n$ are literals, and $\tilde{\forall}$ stands for a bloc of quantifiers over all free variables of the formulae in brackets (universal closure), is called a *u-production*. A *variant of u-production* R is

$$\mathrm{R}\theta \rightleftharpoons \tilde{\forall}(\mathrm{A}_1\theta \wedge \cdots \mathrm{A}_m\theta \Leftarrow \mathrm{B}_1\theta \wedge \cdots \wedge \mathrm{B}_n\theta),$$

where θ is an arbitrary one-to-one correspondence over the set of variables. Let $\mathtt{Prod}$ be the set of all u-productions.

For example in section X presented the following u-production that was discovered by the learning model:

> IF TOTAL number of calcifications is more than 30, AND VOLUME is more than 5 cm^3, AND DENSITY of calcifications is moderate, THEN Malignant.

Let $\mathtt{Fact}_v \subset \mathbf{A}_\mathfrak{L}$ be a set of atoms from A that are valid for verification in algebraic system $\mathfrak{B}$ appearing in practice. Our aim is to investigate as much "extra" facts about $\mathfrak{L}$ as possible, i.e., to predict or explain them. A natural assumption is that we can verify (falsify) each element of

$$\mathtt{Fact}_o \rightleftharpoons \left\{ \mathrm{A}\theta \mid \theta \in \Theta^0, \ \mathrm{A} \in \mathtt{Fact}_v \right\}.$$

Certainly we may postulate our ability to check any literal of $\mathtt{Fact}_v^* \rightleftharpoons \mathtt{Fact}_v \cup \{\neg A \mid A \in \mathtt{Fact}_v\}$. For the rest of the literals (and their conjunctions) the machinery of *probabilistic prediction* will be defined later on. Note that $\mathtt{Fact}_o^* = \mathtt{Fact}_o \cup \mathtt{Fact}_o^\neg$, where $\mathtt{Fact}_o^\neg \rightleftharpoons \{\neg A \mid A \in \mathtt{Fact}_o\}$ is *the complete set of alternatives* allowing a real test.

The *data* are defined as a maximal (logically) consistent subset of the complete set of alternatives, i.e., being given a mapping $\zeta_\mathfrak{B} : \mathtt{Fact}_o \mapsto \{\bot, \top\}$ (here $\bot$ – *"false"*, $\top$ – *"true"*) we conclude that

$$\mathtt{Data}\,[\mathfrak{B}] \rightleftharpoons \{A \mid A \in \mathtt{Fact}_o \text{ and } \zeta_\mathfrak{B}(A) = \top\} \cup$$

$$\{\neg A \mid A \in \mathtt{Fact}_o \text{ and } \zeta_\mathfrak{B}(A) = \bot\}.$$

III. Generality relation between theories

The idea of a *generality relation* between theories can be viewed, for example, as a reduction of the set of properties predicted by the use of these theories. A more general theory (potentially) predicts a greater number of formal features. We start with a generality relation between one-element specifications, i.e., between u-productions.

Definition. For two productions $R_1 \equiv \tilde{\forall}(A_1 \wedge \cdots \wedge A_{m_1} \Leftarrow B_1 \wedge \cdots \wedge B_{n_1})$ and $R_2 \equiv \tilde{\forall}(C_1 \wedge \cdots \wedge C_{m_2} \Leftarrow D_1 \wedge \cdots \wedge D_{n_2})$ a relation $R_1 \succ R_2$ (*"more general than"*) takes place if and only if there exists $\theta \in \Theta$ such that $\{B_1\theta, \cdots, B_{n_1}\theta\} \subseteq \{D_1, \cdots, D_{n_2}\}$, $\{A_1\theta, \cdots, A_{m_1}\theta\} \supseteq \{C_1, \cdots, C_{m_2}\}$, and $n_1 \leqslant n_2$, $m_1 \geqslant m_2$, $\nvdash R_1 \equiv R_2$.

The inclusion of the sets of premises designates that the more general u-production is, then the wider its field of application. The inverse inclusion (for conclusions) says that R_1 predicts a greater number of properties using a smaller premise.

Let $S \subseteq \mathtt{Prod}$. Denote $\mathtt{Fact}\,[S; \mathfrak{B}]$ the set of all $A \in \mathbf{L}_\mathfrak{L}^0$ such that for some $R \in S$ and $\theta \in \Theta^0$,

$$R\theta \equiv (A_1 \wedge \cdots \wedge A_m \Leftarrow B_1 \wedge \cdots \wedge B_n), \text{ holds}$$
$$\{B_1, \cdots, B_n\} \subseteq \mathtt{Data}\,[\mathfrak{B}] \text{ and } A \in \{A_1, \cdots, A_m\}.$$

Thus, $\mathtt{Fact}\,[S; \mathfrak{B}]$ is the set of ground literals predicted according to available data (about the model $\mathfrak{B}$) together with u-productions in S.

In the sequel let $\succcurlyeq$ be a reflexive closure of $\succ$. One should pay attention to the fact: $R_1 \succcurlyeq R_2$ entails that $\mathtt{Fact}\,[\{R_1\}; \mathfrak{B}]$ contains $\mathtt{Fact}\,[\{R_2\}; \mathfrak{B}]$.

Thereafter it isn't difficult to extend the domain of our generality relation to subsets of $\mathtt{Prod}$.

Definition. Let $S, S' \subseteq \mathtt{Prod}$, and for any $R' \in S'$ we find $R \in S$ such that $R \succcurlyeq R'$. In this case we say 'S is not less general than S'' ($S \rhd S'$).

It's straightforward to notice that $\mathtt{Fact}\,[S'; \mathfrak{B}] \subseteq \mathtt{Fact}\,[S; \mathfrak{B}]$ for S and S' from the definition above. Remark that S may include u-productions apart from those, which are generalizations of elements of S'.

IV. Probability/degree of belief

The topic of distributing probability over formulas of propositional logic (as well as over ground statements in a first order language) being widely discussed in a literature and meets Kolmogorov's understanding of probability measure [11]. The following definition is given on the basis of analysis cited in [10].

Definition. *A probability* over $F \subseteq \mathbf{S}_\mathfrak{L}^0$ closed with respect to $\wedge$, $\vee$ and $\neg$, is a function $\mu : F \mapsto [0, 1]$ satisfying the following conditions:

1) if $\vdash \Phi$ ("Φ is a tautology"), then $\mu(\Phi) = 1$;
2) if $\vdash \neg(\Phi \wedge \Psi)$, then $\mu(\Phi \vee \Psi) = \mu(\Phi) + \mu(\Psi)$.

For any ground instance of a u-production its probability is defined as conditional, i.e.,

$$\mu(A_1 \wedge \cdots A_m \Leftarrow B_1 \wedge \cdots \wedge B_n) =$$
$$= \mu(A_1 \wedge \cdots A_m \mid B_1 \wedge \cdots \wedge B_n) =$$
$$\frac{\mu(A_1 \wedge \cdots A_m \wedge B_1 \wedge \cdots \wedge B_n)}{\mu(B_1 \wedge \cdots \wedge B_n)}$$

Let $R \in \mathtt{Prod}$. Denote as $\mathtt{Sub}\,[R]^\mu$ those substitutions $\theta \in \Theta^0$, for which the premise of u-production $R\theta$ has a non-zero probability.

$$\mathtt{Prod}^\mu \rightleftharpoons \{R \in \mathtt{Prod} \mid \mathtt{Sub}\,[R]^\mu \neq \varnothing\};$$
$$\underline{\mu}(R) \rightleftharpoons \inf\{\mu(R\theta) \mid \theta \in \mathtt{Sub}\,[R]^\mu\}, \text{ where } R \in \mathtt{Prod}^\mu.$$

A value of conditional probability serves to characterize our *degree of belief* (and responsible for an *uncertainty relation*) in reliability of different causal connections included in temporary specification. Note that two productions are not necessary comparable with respect to generality relation $\succcurlyeq$; moreover, their premises may not be contained in the complete set of alternatives (and so these productions will be not valid for a direct check in a real structure $\mathfrak{B}$).

V. Simplicity of probabilistic theories

Adding comparison of lower probabilistic estimations to the definition of generality relation we obtain the following definition.

Definition. Let $S, S' \subseteq \mathtt{Prod}^\mu$. We say that S is *more μ-general than* S' iff for every $C' \in S'$ there exists $C \in S$ such that $C \succcurlyeq C'$ and $\underline{\mu}(C) \geqslant \underline{\mu}(C')$, and in at least one of the cases the strong relation $\succ$ takes place.

Hence, μ-generalization allows us to define a more general set S in such a way that the lower estimations of probabilities is not declined. When our belief to the elements of S is no less than that of S', then we have a *simplicity relation* – the set S is simpler than S' in order to describe/predict the properties.

VI. Similarity measure with the empirical content

By elaboration of u-productions we mean the gain of its conditional probability.

Definition. A relation $R_1 \sqsubset R_2$ (*'probabilistic inference'*) for $R_1, R_2 \in \mathtt{Prod}^\mu$ means that $R_1 \succ R_2$ and $\underline{\mu}(R_1) < \underline{\mu}(R_2)$.

Definition. Let π be some requirements to be applied to elements of $\mathtt{Prod}^\mu$, i.e. $\pi : \mathtt{Prod}^\mu \mapsto \{\bot, \top\}$ (value is equal to $\bot$, if u-production satisfies π, and $\top$ – otherwise); $\Pi \leftrightharpoons \{R \in \mathtt{Prod}^\mu \mid \pi(R) = \top\}$. We say that $R_2 \in \Pi$ is a *minimal follower of* $R_1 \in \mathtt{Prod}^\mu$ *relative to* $\sqsubset$ *in* Π (denoted as $R_1 \sqsubset_\pi R_2$), iff $R_1 \sqsubset R_2$ and there is no intermediate u-production $R_{3/2} \in \Pi$ such that $R_1 \sqsubset R_{3/2} \sqsubset R_2$.

In the prediction of a literal H the *similarity measure* for u-productions, which are valid for verification and applicable to the goal H, is equal to conditional probability $\underline{\mu}(\cdot)$. Thus we deal with a uniform measure of similarity.

VII. Learning operator

Definition. A production

$$R \equiv \tilde{\forall}(A_1 \wedge \cdots \wedge A_m \leftarrow B_1 \wedge \cdots \wedge B_n)$$

is called a *maximal specific u-production (ums-production) for prediction of a conjunction* $H \equiv (H_1 \wedge \cdots \wedge H_k)$, where $\{H_1, \cdots, H_k\} \subset \mathbf{L}_\mathfrak{L}$ and $m \leqslant k$, iff the following conditions are satisfied:

1) there is a substitution θ (not necessary ground) such that $\{A_1, \cdots, A_m\} \subseteq \{H_1\theta, \cdots, H_k\theta\}$, $\{B_1, \ldots, B_n\} \subseteq \{B\theta \mid B \in \mathtt{Fact}_v^*\}$;
2) if $D \in \{A_1, \cdots, A_m\}$ and $\theta_o \in \mathtt{Sub}[R]^\mu$, then
$\mu(A_1\theta_o \wedge \cdots \wedge A_m\theta_o) <$
$\mu(A_1\theta_o \wedge \cdots \wedge A_m\theta_o \mid B_1\theta_o \wedge \cdots \wedge B_n\theta_o)$
and $\mu(D\theta_o) < \mu(D\theta_o \mid B_1\theta_o \wedge \cdots \wedge B_n\theta_o)$;
3) there is no $R' \in \mathtt{Prod}^\mu$, for which points (1–2) are hold along with $R \sqsubset R'$;
4) the u-production R can't be generalized up to some $R' \in \mathtt{Prod}^\mu$ satisfying all the previous points (1–3) without decreasing its estimation $\underline{\mu}(\cdot)$.

The conditions above (for corresponding ums-productions) are denoted as 'point.i', $1 \leqslant i \leqslant 4$.

Remark. Though condition point.4 emphasizes the nature of definition, but it isn't necessary for indication. Indeed, if R may be generalized up to R' under preserving point.1–3, then $\underline{\mu}(R) \leqslant \underline{\mu}(R')$ (otherwise we get $R' \sqsubset R$ – that contradicts point.3 for R.

Let $\pi(R) = \top$ be fulfilled for $R \in \mathtt{Prod}^\mu$ iff conditions points.1–2 are satisfied for R and H (the last one is fixed from this moment); denote $\Pi \leftrightharpoons \pi^{-1}(\top)$.

Define the *probabilistic fix-point operator* $T_\pi : 2^{\mathtt{Prod}^\mu} \mapsto 2^{\mathtt{Prod}^\mu}$ as follows: for a set $S \subseteq \mathtt{Prod}^\mu$ it produces

$$S' \leftrightharpoons \{R' \mid R \sqsubset_\pi R' \text{ for some } R \in S\} \cup$$
$$\cup \{R \mid R \in S \cap \Pi \text{ and there is no } R' \text{ such that } R \sqsubset_\pi R'\}.$$

Therefore the operator $T_\pi : S \mapsto S'$ possess important properties:

1) the set S' is always more precise than S (relative to $\succcurlyeq$);

2) the conditional probabilities $\underline{\mu}(\cdot)$ increase during the conversion to more particular cases (and so fuzziness decreases);
3) the similarity measure with the empirical content becomes greater for at least one u-production (in S) when the operator converts S to S' (if not $S = S'$, of course);

As a result the operator T_π is the *enhancement, or learning, operator* in the sense of [1].

Definition. A fix-point (f.p., for short) S of T_π is *optimal* iff there is no other f.p. S' of considered operator, which is more μ-general than S.

Statement. *A subset $S \subseteq \mathtt{Prod}^\mu$ is a fix-point of the operator T_π iff every element of S satisfies points.1–3 for H.*

Corrolary. *A subset $S \subseteq \mathtt{Prod}^\mu$ is an optimal fix-point of the operator T_π iff every element of S is a ums-production for prediction of H.*

Ums-productions may be viewed as a result of performing generalized scheme of the *semantic probabilistic inference* [4], [5], which is realized by the fix-point operator described above. The program system 'Discovery' (see [16], [17], [9], [21]) was developed: it carries out the propositional version of the probabilistic fix-point (learning) operator and was successfully applied to solving many practical tasks [21].

VIII. Extraction of the expert model of breast cancer diagnostic decisions

We applied our method to approximation of the expert model of breast cancer diagnostic decisions that was obtained from the radiologist J.Ruiz [17]. At first we extract this model from the expert by the special procedure using monotone boolean functions [17] and then apply the program system 'Discovery' [16] to approximate this model.

A. Hierarchical Approach

At first we ask an expert to describe particular cases using the binary features. Then we ask a radiologist to evaluate a particular cases, when features take on specific values. A typical query will have the following format: "If feature 1 has value v_1, feature 2 has value v_2, ..., feature n has value v_n, then is a case suspicious of cancer or not?"

Each set of values $(v_1, v_2, ..., v_n)$ represent a possible clinical case. It is practically impossible to ask a radiologist to generate diagnosis for thousands of possible cases. A hierarchical approach combined with the use of the property of monotonicity makes the problem manageable. We construct a hierarchy of medically interpretable features from a very generalized level to a less generalized level. This hierarchy follows from the definition of the 11 medically oriented binary attributes. The medical expert indicate that the original 11 binary attributes $w_1, w_2, w_3, y_1, y_2, y_3, y_4, y_5, x_3, x_4, x_5$ could be organized in terms of a hierarchy with development of two new generalized attributes x_1, depending on attributes w_1, w_2, w_3, and x_2, depending on attributes y_1, y_2, y_3, y_4, y_5.

A new generalized feature, x_1 – 'Amount and volume of calcifications' with grades (0 - 'benign' and 1 - 'cancer') was introduced based on features: w_1 – number of calcifications/cm3, w_2 – volume of calcification, cm3 and w_3 – total number of calcifications. We view x_1 as a function $g(w_1, w_2, w_3)$ to be identified. Similarly a new feature x_2 – 'Shape and density of calcification' with grades: (1) for 'cancer' and (0)-'benign' generalizes features: y_1 – 'irregularity in shape of individual calcifications' y_2 – 'variation in shape of calcifications' y_3 – 'variation in size of calcifications' y_4 – 'variation in density of calcifications' y_5 – 'density of calcifications'. We view x_2 as a function $x_2 = h(y_1, y_2, y_3, y_4, y_5)$ to be identified for cancer diagnosis.

As result we have a decomposition of our task as follows:

$$f(x_1, x_2, x_3, x_4, x_5) =$$

$$f(g(w_1, w_2, w_3), h(y_1, y_2, y_3, y_4, y_5), x_3, x_4, x_5).$$

B. Monotonicity

Giving the above definitions we can represent clinical cases in terms of binary vectors with five generalized features as: $(x_1, x_2, x_3, x_4, x_5)$. Let us consider two clinical cases that are represented by the two binary sequences: (10110) and (10100). If radiologist correctly diagnose (10100) as cancer, then, by utilizing the property of monotonicity, we can also conclude that the clinical case (10110) should also be cancer. Medical expert agreed with presupposition about monotonicity of the functions $f(x_1, x_2, x_3, x_4, x_5)$ and $h(y_1, y_2, y_3, y_4, y_5)$.

Let us describe the interview with an expert using minimal sequence of questions to completely infer a diagnostic function using monotonicity. This sequence is based on fundamental Hansel lemma [15]. We omit a detailed description of the specific mathematical steps. They can be found in [18]. Table 1 illustrates this.

C. Expert model extraction

Columns 2 and 3 present values of above defined functions f and h. We omit a restoration of function $g(w_1, w_2, w_3)$ because few questions are needed to restore this function. All 32 possible cases with five binary features $\langle x_1, x_2, x_3, x_4, x_5 \rangle$ are presented in column 1 in table 1. They are grouped and the groups are called Hansel chains [17]. The sequence of chains begins with the shortest chain 1 – (01100) < (11100) for five binary features. Then largest chain 10 consists of 6 ordered cases: (00000) < (00001) < (00011) < (00111) < (01111) < (11111). The chains are numbered there from 1 to 10 and each case has its number in the chain, e.g., 1.2 means the second case in the first chain. Asterisks in columns 2 and 3 mark answers obtained from an expert, e.g., 1* for case (01100) in column 3 means that the expert answered 'yes'. The answers for some other chains in column 3 are automatically obtained using monotonicity. The value f(01100) = 1 for case 1.1 is extended for cases 1.2, 6.3. and 7.3 in this way. Similarly values of the monotone Boolean functions h are computed using the table 1. The attributes in the sequence (10010)

are interpreted as y_1, y_2, y_3, y_4, y_5 for the function h instead of x_1, x_2, x_3, x_4, x_5. The Hansel chains are the same if the number of attributes is the same five in this case.

Column 5 and 6 list cases for extending functions' values without asking an expert. Column 5 is for extending functions' values from 1 to 1 and column 6 is for extending them from 0 to 0. If an expert gave an answer opposite (f(01100) = 0) to that presented in table 1 for function f in the case 1.1, then this 0 value could be extended in column 2 for cases 7.1 (00100) and 8.1 (01000). These cases are listed in column 5 for case (01100). There is no need to ask an expert about cases 7.1 (00100) and 8.1 (01000). Monotonicity provides the answer. The negative answer f(01100) = 0 can not be extended for f(11100). An expert should be queried regarding f(11100). If his/her answer is negative f(11100) = 0 then this value can be extended for cases 5.1. and 3.1 listed in column 5 for case 1.2. Relying on monotonicity, the value of f for them will also be 0.

The total number of cases with asterisk (*) in columns 2 and 3 are equal to 13 and 12. These numbers show that 13 questions are needed to restore the function $f(x_1, x_2, x_3, x_4, x_5)$ and 12 questions are needed to restore the function $h(y_1, y_2, y_3, y_4, y_5)$. This is only 37.5% of 32 possible questions. The full number of questions for the expert without monotonicity and hierarchy is $2^{11} = 2048$.

IX. APPROXIMATION OF THE EXPERT MODEL BY LEARNING OPERATOR

For the Approximation of the expert model we used the program system 'Discovery' [16], that realizes the propositional case of the probabilistic fix-point learning operator. We discovered several dozens diagnostic rules that were statistically significant on the 0.01, 0.05 and 0.1 levels of (F-criterion). Rules were extracted using 156 cases (73 malignant, 77 benign, 2 highly suspicious and 4 with mixed diagnosis). In the Round-Robin test our rules diagnosed 134 cases and refused to diagnose 22 cases. The total accuracy of diagnosis is 86%. Incorrect diagnoses were obtained in 19 cases (14% of diagnosed cases). The false-negative rate was 5.2% (7 malignant cases were diagnosed as benign) and the false-positive rate was 8.9% (12 benign cases were diagnosed as malignant). Some of the rules are shown in table 2. This table presents examples of discovered rules with their statistical significance. In this table:

- 'NUM' – number of calcifications per cm^3;
- 'VOL' – volume in cm^3;
- 'TOT' – total number of calcifications;
- 'DEN' – density of calcifications;
- 'VAR' – variation in shape of calcifications;
- 'SIZE' – variation in size of calcifications;
- 'IRR' – irregularity in shape of calcifications;
- 'SHAPE' – shape of calcifications.

We studied three levels of similarity measure: 0.7, 0.85 and 0.95. A higher level of conditional probability decreases the number of rules and diagnosed patients, but increases accuracy of diagnosis.

	1	2	3	4	5	6	7
Table 1. Dynamic sequence of questions to expert							
Number	f	h	Monotonic extrapolation		Chain	Case
	Diagnose	Form and V	$1 \mapsto 1$	$0 \mapsto 0$		
(01100)	1*	1*	1.2, 6.3, 7.3	7.1, 8.1	Chain 1	1.1
(11100)	1	1	6.4, 7.4	5.1, 3.1		1.2
(01010)	0*	1*	2.2, 6.3, 8.3	6.1, 8.1	Chain 2	2.1
(11010)	1*	1	6.4, 8.4	3.1, 6.1		2.2
(11000)	1*	1*	3.2	8.1, 9.1	Chain 3	3.1
(11001)	1	1	7.4, 8.4	8.2, 9.2		3.2
(10010)	0*	1*	4.2, 9.3	6.1, 9.1	Chain 4	4.1
(10110)	1*	1	6.4, 9.4	6.2, 5.1		4.2
(10100)	1*	1*	5.2	7.1, 9.1	Chain 5	5.1
(10101)	1	1	7.4, 9.4	7.2, 9.2		5.2
(00010)	0	0*	6.2, 10.3	10.1	Chain 6	6.1
(00110)	1*	0*	6.3, 10.4	7.1		6.2
(01110)	1	1	6.4, 10.5			6.3
(11110)	1	1	10.6			6.4
(00100)	1*	0*	7.2, 10.4	10.1	Chain 7	7.1
(00101)	1	0*	7.3, 10.4	10.2		7.2
(01101)	1	1*	7.4, 10.5	8.2, 10.2		7.3
(11101)	1	1	5.6			7.4
(01000)	0	1*	8.2	10.1	Chain 8	8.1
(01001)	1*	1	8.3	10.2		8.2
(01011)	1	1	8.4	10.3		8.3
(11011)	1	1	10.6	9.3		8.4
(10000)	0	1*	9.2	10.1	Chain 9	9.1
(10001)	1*	1	9.3	10.2		9.2
(10011)	1	1	9.4	10.3		9.3
(10111)	1	1	10.6	10.4		9.4
(00000)	0	0	10.2		Chain 10	10.1
(00001)	0*	0	10.3			10.2
(00011)	1*	0	10.4			10.3
(00111)	1	1*	10.5			10.4
(01111)	1	1	10.6			10.5
(11111)	1	1				10.6
Questions	13	12				

Table 2. Examples of discovered diagnostic rules						
Diagnosis rule	f-criteria		Value. f-criteria			Precision on control
			0.01	0.05	0.1	
If $10 < \text{NUM} < 20$ and $\text{VOL} > 5$ then *malignant*	NUM	0.0029	+	+	+	93.3%
	VOL	0.0040	+	+	+	
If $\text{TOT} > 30$ and $\text{VOL} > 5$ and DEN is *moderate* then *malignant*	TOT	0.0229	-	+	+	100.0%
	VOL	0.0124	-	+	+	
	DEN	0.0325	-	+	+	
If VAR is *marked* and $10 < \text{NUM} < 20$ and IRR is *moderate* then *malignant*	VAR	0.0044	+	+	+	100.0%
	NUM	0.0039	+	+	+	
	IRR	0.0254	-	+	+	
If SIZE is *moderate* and SHAPE is *mild* and IRR is *mild* then *benign*	SIZE	0.0150	-	+	+	92.86%
	SHAPE	0.0114	-	+	+	
	IRR	0.0878	-	-	+	

Results for them are marked as Discovery1, Discovery2 and Discovery3. We extracted 44 statistically significant diagnostic rules for 0.05 level of F -criterion with a conditional probability no less than 0.75 (Discovery1). There were 30 rules with a conditional probability no less than 0.85 (Discovery2) and 18 rules with a conditional probability no less than 0.95 (Discovery3). The most reliable 30 rules delivered a total accuracy of 90%, and the 18 most reliable rules performed with 96.6% accuracy with only 3 false positive cases (3.4%).

X. Decision rule (model) extracted from the expert through monotone Boolean functions

We obtained Boolean expressions for function $h\,(y_1, y_2, y_3, y_4, y_5)$ ('shape and density of calcification') from the information depicted in table 1 with the following steps:

- Find all the maximal lower units for all chains as elementary conjunctions;
- Take the disjunction of obtained conjunctions;
- Exclude the redundant terms (conjunctions) from the end formula.

Using 1 and 3 columns we have

$$x_2 = h\,(y_1, y_2, y_3, y_4, y_5) = y_2 y_3 \vee y_2 y_4 \vee y_1 y_2 \vee y_1 y_4 \vee y_1 y_3 \vee$$

$$\vee y_2 y_3 y_5 \vee y_2 \vee y_1 \vee y_3 y_4 y_5 \equiv y_2 \vee y_1 \vee y_3 y_4 y_5.$$

Function $g\,(w_1, w_2, w_3) = w_2 \vee w_1 w_3$ we may obtain by direct $2^3 = 8$ questions for the expert.
Using 1 and 2 columns we have

$$f\,(\overline{x}) = x_2 x_3 \vee x_1 x_2 x_4 \vee x_1 x_2 \vee x_1 x_3 x_4 \vee x_1 x_3 \vee x_3 x_4 \vee x_3$$

$$\vee x_2 x_5 \vee x_1 x_5 \vee x_4 x_5 \equiv x_1 x_2 \vee x_3 \vee (x_2 x_1 x_4)\, x_5 \equiv$$

$$(w_2 \vee w_1 w_3)\,(y_1 \vee y_2 \vee y_3 y_4 y_5) \vee x_3 \vee$$

$$\vee\,(y_1 \vee y_2 \vee y_3 y_4 y_5)\,(w_2 \vee w_1 w_3)\, x_4 x_5.$$

XI. Comparison of the expert model with its approximation by learning operator

For compare rules discovered by the learning operator (Discovery system) with the expert model we asked the expert to evaluate this rules. Below we present some rules, discovered by Discovery system, and radiologists comments regarding these rules as approximation of his model.

> IF TOTAL number of calcifications is more than 30, AND VOLUME is more than 5 cm^3, AND DENSITY of calcifications is moderate, THEN Malignant.

f-criterion significant for 0.05. Accuracy of diagnosis for test cases – 100%. Radiologist's comment - this rule might have promise, but I would consider it risky.

> IF VARIATION in shape of calcifications is marked, AND NUMBER of calcifications is between 10 and 20, AND IRREGULARITY in shape of calcifications is moderate, THEN Malignant.

f-criterion significant for 0.05. Accuracy of diagnosis for test cases – 100%. Radiologist's comment - I would trust this rule.

> IF variation in SIZE of calcifications is moderate, AND variation in SHAPE of calcifications is mild, AND IRREGULARITY in shape of calcifications is mild, THEN Benign.

f-criterion significant for 0.05. Accuracy of diagnosis for test cases – 92.86%. Radiologist's comment - I would trust this rule.

References

[1] Kovalerchuk B. Ya., Perlovsky L. I. *Dynamic logic of phenomena and cognition*. IJCNN, 2008, pp. 3530–3537.
[2] Perlovsky L. I. *Toward physics of the mind: concepts, emotions, consciousness, and symbols* // Physics of Life Reviews, 3, 2006, pp. 23–55.
[3] Perlovsky L. I. *Neural Networks, Fuzzy Models and Dynamic Logic* // R. Kohler and A. Mehler, eds., Aspects of Automatic Text Analysis (Festschrift in Honor of Burghard Rieger), Springer, Germany, 2007, pp. 363-386.
[4] Vityaev E. E. *The logic of prediction* // Mathematical Logic in Asia 2005, Proceedings of the 9th Asian Logic Conference, eds. Goncharov S.S., Downey R. and Ono H., August 16–19, Novosibirsk, Russia, World Scientific Publisher, 2006, pp. 263–276.
[5] Smerdov S. O., Vityaev E. E. *Probability, logic & learning synthesis: formalizing prediction concept* // Siberian Electronic Mathematical Reports, vol. 9, 2009, pp. 340–365., in russian, english abstract.
[6] Vityaev E. E., Smerdov S. O. *New definition of prediction without logical inference* // Proceedings of the IASTED international conference on Computational Intelligence (CI 2009), ed. Kovalerchuk B. Ya., August 17-19, Honolulu, Hawaii, USA, pp. 48–54.
[7] Evgenii Vityaev, Principals of brain activity, supported the functional systems theory by P.K. Anokhin and emotional theory by P.V. Simonov, *Neiroifformatics*, v.3, N1, 2008, pp. 25-78
[8] Akexander Demin, Evgenii Vityaev, Logical model of adaptive control system. *Neiroifformatics*, v.3, N1, 2008, pp. 79-107
[9] Evgenii Vityaev. Knowledge discovery. Computational cognition. Cognitive processes modeling. Novosibirsk State University, Novosibirsk, 2006. pp.293.
[10] Halpern J. Y. *An analysis of first-order logics of probability*. In: Artificial Intelligence, 46, 1990, pp. 311–350.
[11] Shiryaev A. N. *Probability*. Springer, 1995.
[12] Keisler H. J., Chang C. C. *Model theory*. Elsevier, 1990.
[13] Maltsev A. I. *Algebraic systems*. Springer-Verlag, 1973.
[14] Lloyd J. W. *Foundations of logic programming*. Springer-Verlag, 1987.
[15] Hansel G. Sur le nombre des fonctions Boolenes monotones den variables // C. R. Acad. Sci. Paris, vol. 262, 20, 1966, pp. 1088–1090.
[16] Kovalerchuk B. Ya., Vityaev E. E. *Data mining in finance: advances in relational and hybrid methods*. Kluwer Academic Publisher, 2000.
[17] Kovalerchuk B. Ya., Vityaev E. E., Ruiz J. F. *Consistent and complete data and "expert" mining in medicine* // Medical data mining and knowledge discovery, Springer, 2001, pp. 238–280.
[18] Kovalerchuk B, Talianski V. Comparison of empirical and computed fuzzy values of conjunction. Fuzzy Sets and Systems 46: 49-53, 1992.
[19] The Probabilistic Mind. Prospects for Bayesian cognitive sciense // Eds. Nick Chater, Mike Oaksford, Oxfor University Press, 2008, pp.536
[20] Probabilistic models of cognition // Special issue of the journal *Trends in cognitive science*, v.10, Issue 7, 2006, pp. 287-344
[21] Scientific Discovery website. http://math.nsc.ru/AP/ScientificDiscovery